

令和 4 年 6 月 3 日現在

機関番号：34315

研究種目：基盤研究(C) (一般)

研究期間：2019～2021

課題番号：19K12226

研究課題名(和文)顕微鏡画像データに基づく因果構造マイニング法の開発

研究課題名(英文)Development of Computational Methods for Mining Causal Relationships Based on Microscopic Images

研究代表者

遠里 由佳子 (Tohsato, Yukako)

立命館大学・情報理工学部・教授

研究者番号：80346171

交付決定額(研究期間全体)：(直接経費) 3,400,000円

研究成果の概要(和文)：タイムラプス顕微鏡画像データから得られる表現型などの因果関係を計算機で推論する時系列因果推論手法の確立を目指した。非線形な時系列の因果推論の従来法として、Convergent Cross Mapping (CCM)やGranger因果性テストのノンパラメトリックな拡張(以下、NPMR)に着目した。そして、NPMRを基盤に、新しい時系列因果推論の手法を提案した。人工的に生成した時系列で、CCMやNPMRと比較し、提案手法の有効性を確認した。微小管をGFPで可視化した線虫を蛍光顕微鏡で撮影したタイムラプス画像データを用いて、雌性前核と精子星状体の動態の生物学的知識の発見を試みた。

研究成果の学術的意義や社会的意義

ライブセルイメージング技術の普及に伴い、大量のタイムラプス画像が容易に得られている。そのため、得られる画像から自動的に因果関係を推論する手法の確立は、生命科学研究の効率化や自動化において重要な課題となっている。モデル生物であり、大規模なオープンデータが既に公開されている線虫の顕微鏡画像由来の時系列データから、因果推論を試みることは、生命科学の分野で今後産出される他のモデル生物を対象とした時系列への応用を意味している。因果関係の正解を設定することができる、人工データやシミュレーションデータを用いて、それらの推論の精度比較を行なうことは、提案手法の妥当性を評価する上で確実なステップとなる。

研究成果の概要(英文)：We aimed to establish a computational method for inferring causal relationships such as phenotypes from time-lapse microscopy images data. We focused on Convergent Cross Mapping (CCM) and a nonparametric extension of the Granger causality test (NPMR) as the previous methods targeting for nonlinear time-series data. We then proposed a novel method based on the NPMR. We confirmed the effectiveness of our proposed method by comparing it with CCM and NPMR on artificially generated time series data. Using time-lapse images of *C. elegans* with microtubules visualized with GFP observed by fluorescence microscopy, we tried to obtain new biological knowledge about the dynamics of the female pronucleus and microtubule sperm asters

研究分野：バイオインフォマティクス

キーワード：因果推論 時系列データ解析 表現型解析 *C. elegans* データ駆動型解析 バイオイメージインフォマティクス システム生物学

1. 研究開始当初の背景

生命科学の分野では、大規模・精密な撮影を可能にするライブイメージング技術と、より効率的な遺伝子改変技術、ロボット技術の開発と導入により、大量のタイムラプス画像が生まれつつある。そして、得られた画像データを対象とする、畳み込みニューラルネットワークなどを用いた高精度な画像処理技術の発展により、細胞や組織などの生命現象の時空間情報を数値として含む定量データが得られるようになった。時空間定量データから特徴抽出を経て得られる非線形な時系列データは、生命現象を決定論的かつ非線形なシステムとして認識し解析する上で欠かすことができない。そのため、人が生命現象を解析して、時間変化する複数の変量間の原因と結果の関係(因果関係)を発見する試みが繰り返され試みられている。しかし、そうした顕微鏡画像に由来する時系列データ間の因果関係を計算機でマイニングする試みはほとんど実例がないという問題があった。

非線形な時系列データ間の因果関係の推論手法として、Convergent Cross Mapping (引用文献①; 以下、CCM と省略)や Non-Parametric Multiplicative Regression を用いる手法 (引用文献②; 以下、NPMR と省略)の報告があった。しかし、生命科学の分野で得られる画像データ由来の時系列に対して因果推論を試みた研究は、研究開始当初の時点では報告されていなかった。例えば、CCM は特に生態学の分野で得られる時系列の解析に実績がある。しかし、最低限 40 時点、理想的には 1000 時点の時系列長を必要とすることが報告されており、画像データ由来の時系列に向かない可能性が高かった。一方、NPMR はノイズの多い時系列に適するように提案された時系列因果推論の手法であり、睡眠に関する生体データなどに応用された実績はあったが、生命科学の分野で得られる画像由来の時系列データの解析において、必要とする時系列長など不明な点が多かった。

2. 研究の目的

そのような背景のもと、本研究計画では、データ駆動型サイエンスの加速にむけた必須なアプローチとして、時系列因果推論に着目し、タイムラプス画像データを対象とする因果構造マイニング法の確立を目指した(図 1)。研究目的は、(1) 時系列を作成するための特徴抽出などの前処理における多重共線性や汎用性の検討と、(2) 短くノイズの多い時系列に適した時系列因果推論手法の提案、(3) 生物の実データへの応用からなつた。特に (3) では、タイムラプス蛍光顕微鏡により、線虫の 1 細胞期の雌性前核の時空間動態を観察して得られた画像データに着目し、画像処理と特徴抽出を経て得られる時系列に対して因果推論に挑戦した。

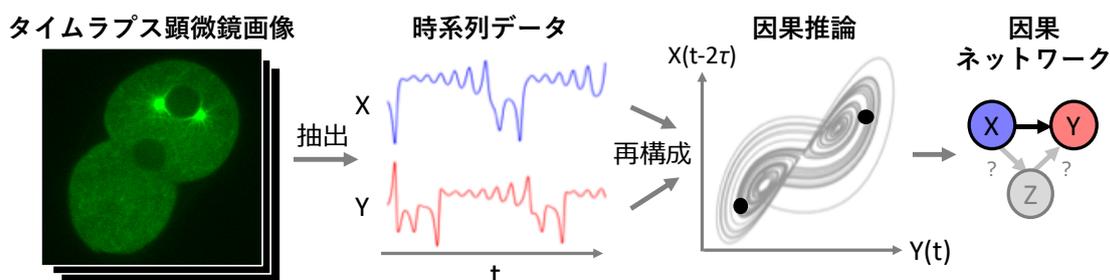


図 1 : 画像由来の因果構造マイニングの各ステップ

3. 研究の方法

(1) では、時系列を生成する際の特徴選択法の多重共線性の評価と、より汎用性のある特徴抽出法として楕円フーリエ記述子と呼ばれる形状解析手法の導入を検討する。(2)の時系列因果推論手法の提案では、因果関係の強さが任意に設定できる結合ロジスティックマップなどで生成した非線形な時系列を用いて、予測精度の評価を行った。(3)では、顕微鏡画像データを画像処理し、雌性前核の移動速度と、精子星状体の面積の間の因果関係の推論を行った。

4. 研究成果

(1)特徴抽出など前処理法の検討

より一般的な解析を実地できるよう、楕円フーリエ記述子を用いた輪郭形状の分析も検証した。楕円フーリエ記述子では、輪郭を楕円フーリエ係数で近似し、係数を対象の回転などに不変となるよう標準化することで、情報の抽出と復元が可能であることを確認した。さらに、相関係数を用いて特徴量間の多重共線性を確認した。また、時系列データや時空間定量データを統一的に格

納できる形式や、その形式にあった可視化技術についての検討も行った。

(2) 時系列因果推論の方法論の提案

因果関係がある非線形な時系列に対する推論精度が、ノイズの有無によりどのように変化するかを、提案手法と NPMR、CCM と比較し検証した。加えて、因果関係がない時系列としてサロゲートデータを作成し、誤検出の割合も確認した。提案手法と従来法である NPMR と CCM の因果推論の精度を、結合ロジスティックマップで生成した時系列長 $N=25, 50, 100, 1000$ の時系列を用いて比較した。ただし、影響率と呼ばれるパラメータを調節し、生成される 2 つの時系列 X と Y が、原因 X 、結果 Y (以下、 $X \rightarrow Y$) となるよう設定した。そして、 $X \rightarrow Y$ が推論できたときに正解とした。代表的な時系列長 $N=25, 50, 100, 1000$ とし (各サンプル数=2000)、時系列長が異なる条件で、提案手法、NPMR、CCM の正解率の違いを確認したところ (表 1)、 $N=25$ の短時系列では、提案手法が最も高い精度 71.0% を達成した。さらに、 $N=25$ の短時系列を対象に、プロセスノイズを加えたときの正解率の変化を確認したところ (図 2)、正解率 60% 以上を達成する上では、提案手法が最も良いことを確認した。サロゲートデータを用いた検出より因果関係がない時系列を誤検出する確率は 10% 程度に抑えられることを確認した。しかし、提案手法は、解析に用いる時系列長が長くなると推論精度が下がることから、改善案も検討している。

表 1：時系列長の変化に伴う正解率の推移 (%)

	時系列長			
	25	50	100	1000
提案手法	71.0	22.4	3.7	0.1
NPMR	2.5	0.1	0.0	0.0
CCM	40.0	47.9	50.9	80.5

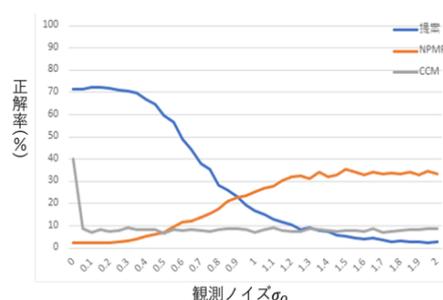


図 2：時系列長 25 の場合のノイズ量と正解率の変化

なお、上記の提案手法のほかにも、CCM に非線形回帰である RBF ネットワークを導入した Cross Map Smoothness (引用文献③) に着目し、サポートベクター回帰への変更などを試みた。従来法である RBF ネットワークを利用した場合と、サポートベクター回帰を利用した場合とでノイズに対する頑強性を予測誤差で評価した。こちらは評価がまとも次第、研究発表する。

(3) 生物の実データへの応用

最初に、対象のタイムラプス画像データに対し、胚領域、雌性前核領域、星状体領域を検出する画像処理のアルゴリズムを開発した (画像処理のアルゴリズムは、二値化に大津法を、ノイズ除去に検出された領域のトラッキングを用いている)。次に、検出領域から A：雌性前核の移動速度と、B：星状体の合計体積を求めた。そして、(条件 1) すべての時系列と、(条件 2) 核の移動速度が活発な時期のみに限定した時系列で、 $A \rightarrow B$ や $B \rightarrow A$ の因果関係を判定した (図 3 として条件 2 のみ抜粋)。その結果、条件 2 で、 $B \rightarrow A$ 、星状体の体積が原因となり雌性前核の移動速度が変化することが示された。条件 1 と 2 の推論結果の違いから、核の移動が活発な時期と、そうではない時期で、別の制御システムが機能していることが示唆された。

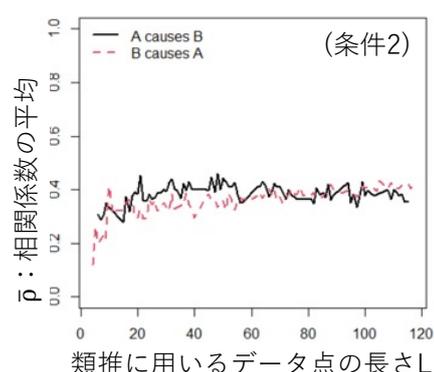


図 3：実データでの推論結果

<引用文献>

- ① Sugihara, G., May, R., Ye, H., Hsieh, C. Deyle, E., Fogarty, M., and Munch, S. (2012) Detecting Causality in Complex Ecosystems, *Science*, vol. 338, no. 6106, pp. 496–500.
- ② Nicolaou, N., and Constandinou, T. G. (2016) A Nonlinear Causality Estimator Based on Non-Parametric Multiplicative Regression, *Frontiers in Neuroinformatics*, vol. 10, pp. 1–21.
- ③ Ma, H., Aihara, K., and Chen, L. (2014) Detecting Causality from Nonlinear Dynamics with Short-term Time Series, *Scientific Reports*, vol. 4, p. 7464.

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Koji Kyoda, Kenneth H. L. Ho, Yukako Tohsato, Hiroya Itoga, Shuichi Onami	4. 巻 15
2. 論文標題 BD5: An open HDF5-based data format to represent quantitative biological dynamics data	5. 発行年 2020年
3. 雑誌名 PLOS ONE	6. 最初と最後の頁 e0237468
掲載論文のDOI（デジタルオブジェクト識別子） 10.1371/journal.pone.0237468	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計6件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 中谷 圭志, 遠里 由佳子
2. 発表標題 線虫の細胞核の時系列データの HDP-HSMM を用いた分節化
3. 学会等名 情報処理学会研究報告
4. 発表年 2020年

1. 発表者名 野崎 弘晃, 遠里 由佳子
2. 発表標題 撮影条件の異なる細胞画像の領域検出におけるデータ拡張の有効性
3. 学会等名 第9回生命医薬情報連合大会 (IIBMP2020)
4. 発表年 2020年

1. 発表者名 中谷 圭志, 遠里 由佳子
2. 発表標題 確率的生成モデルを用いた線虫初期発生の分節化
3. 学会等名 第9回生命医薬情報連合大会 (IIBMP2020)
4. 発表年 2020年

1. 発表者名 大山 鷹志, 遠里 由佳子
2. 発表標題 非線形な時系列に対するガウス過程回帰を用いた因果推論
3. 学会等名 第10回生命医薬情報連合大会(IIBMP2021)
4. 発表年 2021年

1. 発表者名 野崎 弘晃, 遠里 由佳子
2. 発表標題 領域検出における時系列を用いたドメイン適用手法と細胞画像への応用
3. 学会等名 第24回画像の認識・理解シンポジウム(MIRU2021)
4. 発表年 2021年

1. 発表者名 大山 鷹志, 遠里 由佳子
2. 発表標題 代謝ネットワーク解析に向けた予測誤差の状態空間を用いる時系列因果推論法の提案
3. 学会等名 情報処理学会研究報告
4. 発表年 2022年

〔図書〕 計1件

1. 著者名 京田 耕司, ホー ケネス, 糸賀 裕弥, 遠里 由佳子, 大浪 修一	4. 発行年 2020年
2. 出版社 可視化情報学会論文誌 特集号 生物学のデータと可視化	5. 総ページ数 4
3. 書名 生命現象の時空間動態情報の可視化	

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------